Weekly Report(July.22.2019-July.28.2019)

Shanghai Jiao Tong University
1747836307@qq.com
Shaopu Rui

Abstract

This week I continue to train Resnets and get an average accuracy of 95%. And this mini project seems to come to an end.

1 Question from last week

- First, I run more different models to ensure it is not a coincidence. The result are shown at Table 1. The performances are relatively steady and still worse than SGD's. Therefore, the phenomenon I metioned last report still exists. It seems not a coincidence.

Table 1: momentum=0.95, epochs=100

Learning Rate	Batch Size	Accuracy
0.01	256	78.42%
0.003	256	74.21%
0.001	256	72.33%
0.0003	256	63.17%
0.0001	256	47.81%

- The loss curves are shown in Figure 1. obviously when learning rate becomes smaller, model need more epochs to converge. Last report I didn't realize this since models optimized by SGD can all converge in 100 epochs. Sorry about this mistake.
- Give the model larger learning rate and it can indeed converge in 100 epochs with a good performance. The results are shown in Table 2 and loss curves are shown in Figure 2.

Table 2: momentum=0.95, epochs=100

Learning Rate	Batch Size	Accuracy
1	256	87.29%
0.3	256	87.97%
0.1	256	86.58%
0.03	256	83.41%

2 ResNet

Since there are several models in ResNets and train a model once needs much time, I just trained ResNet-18 with different hyper parameters and deeper ResNets with same hyper parameter. Hyper parameters and result of each model are shown as following:

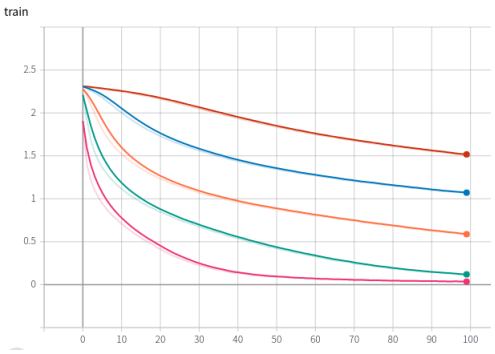

Figure 1: small learning rate

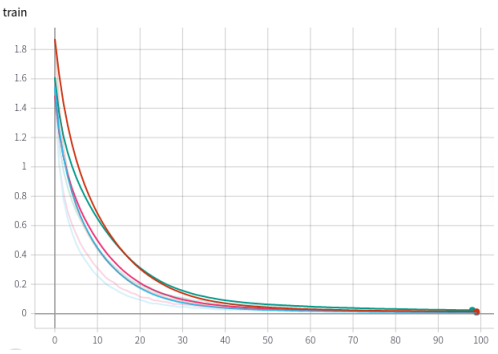

Figure 2: large learning rate

Figure 3: loss curves with Adadelta

- In this part, replace the fc layer for all nets with a new fc layer initialized with **nn.init.normal_(params, mean=0.0, std=0.1)**
- The accuracies of ResNet-18 are shown in table 3. Generally It performs well.
  - As learning rate becomes smaller, net performs better since loss can be more closer to minimum point.
  - Since larger batch size will run out of memory, small batch size is more used here. Within this scale, larger batch size works better.

Table 3: momentum=0.9, epochs=10

Learning Rate	Batch Size	Accuracy
0.01	8	86.18%
0.01	16	91.59%
0.01	32	92.99%
0.003	8	94.93%
0.003	16	95.68%
0.003	32	95.53%
0.001	8	92.15%
0.001	16	95.56%

- ResNet-34, ResNet-50, ResNet-101 are also trained with the hyper parameters performing best in ResNet-18. The Accuracies are shown in table 4 and their loss curves are shown in Figure 4.

Table 4: learning_rate0.003, batch_size=16, momentum=0.9, epochs=20

Net	Accuracy
ResNet-18	95.68%
ResNet-34	95.32%
ResNet-50	95.45%
ResNet-101	96.23%

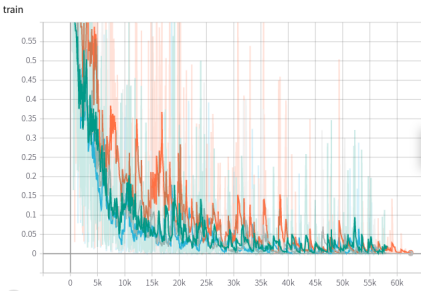

Figure 4: ResNets Loss Curves

3 Summary of Mini Project

Generally Mini project has come to an end. The best performances of these three models are shown as following:

- AlexNet: with learning rate=0.01, batch size=128, momentum=0.9, get accuracy of 88.09%.
- VGG Net: with learning rate=0.0003, batch size=8, momentum=0.9, get accuracy of 91.10%.
- ResNets: ResNet-101 with learning rate=0.003, batch size=16, momentum=0.9, get accuracy of 96.23%.

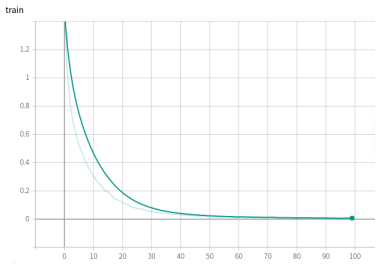

Figure 5: AlexNet Loss Curve

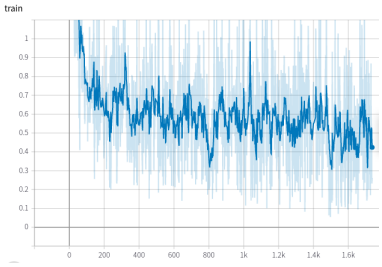

Figure 6: VGG Loss Curve

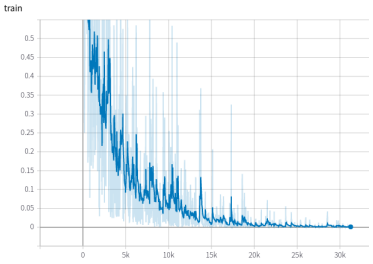

Figure 7: ResNet Loss Curve

Figure 8: loss curves of best performance